[AWS Deployment and Service Models 2](#_Toc201579746)

[AWS Cost Analysis 3](#_Toc201579747)

[AWS Global infrastructure 4](#_Toc201579748)

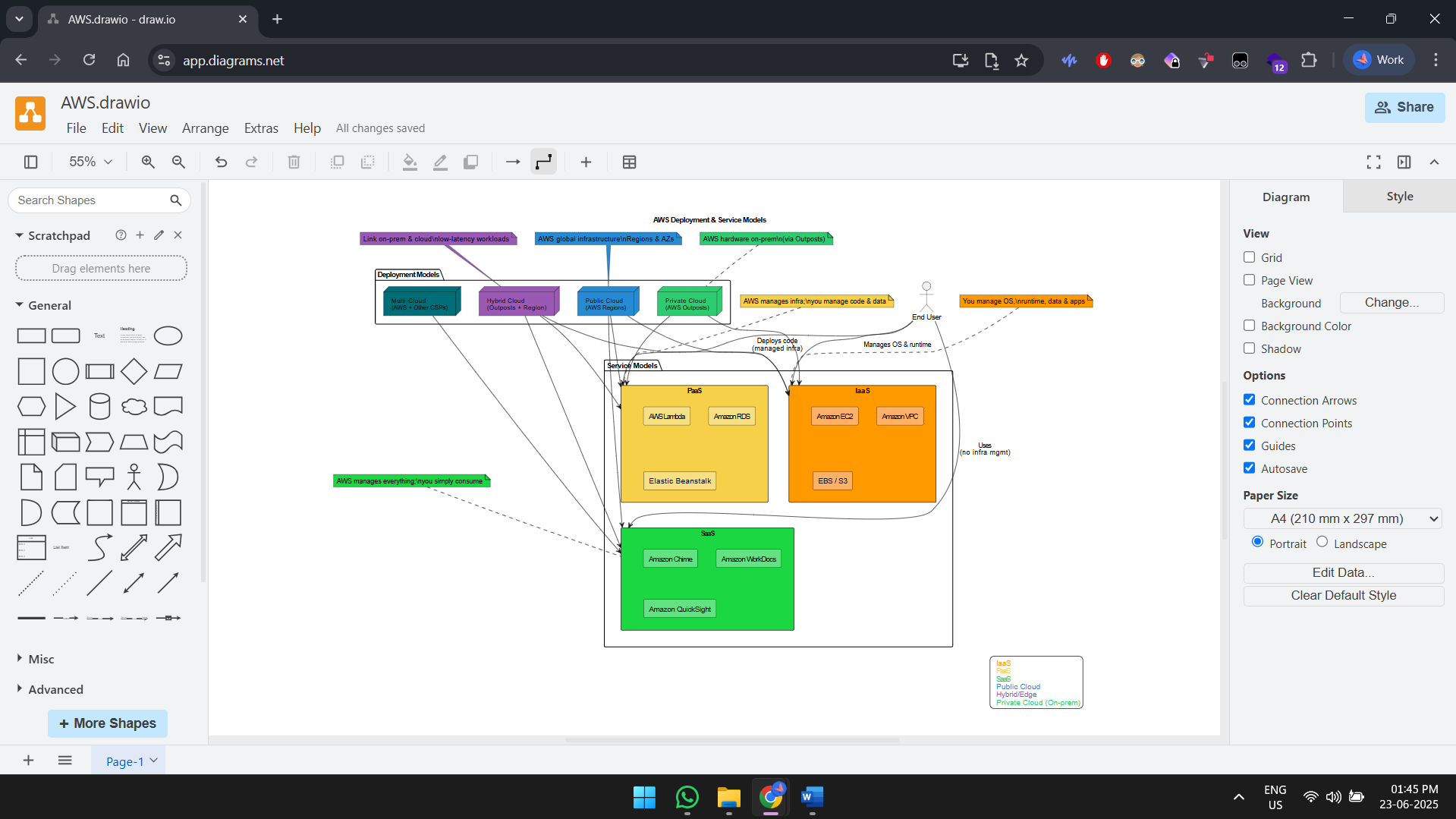
[AWS IAM 5](#_Toc201579749)

[AWS VPC 6](#_Toc201579750)

[AWS Lambda 7](#_Toc201579751)

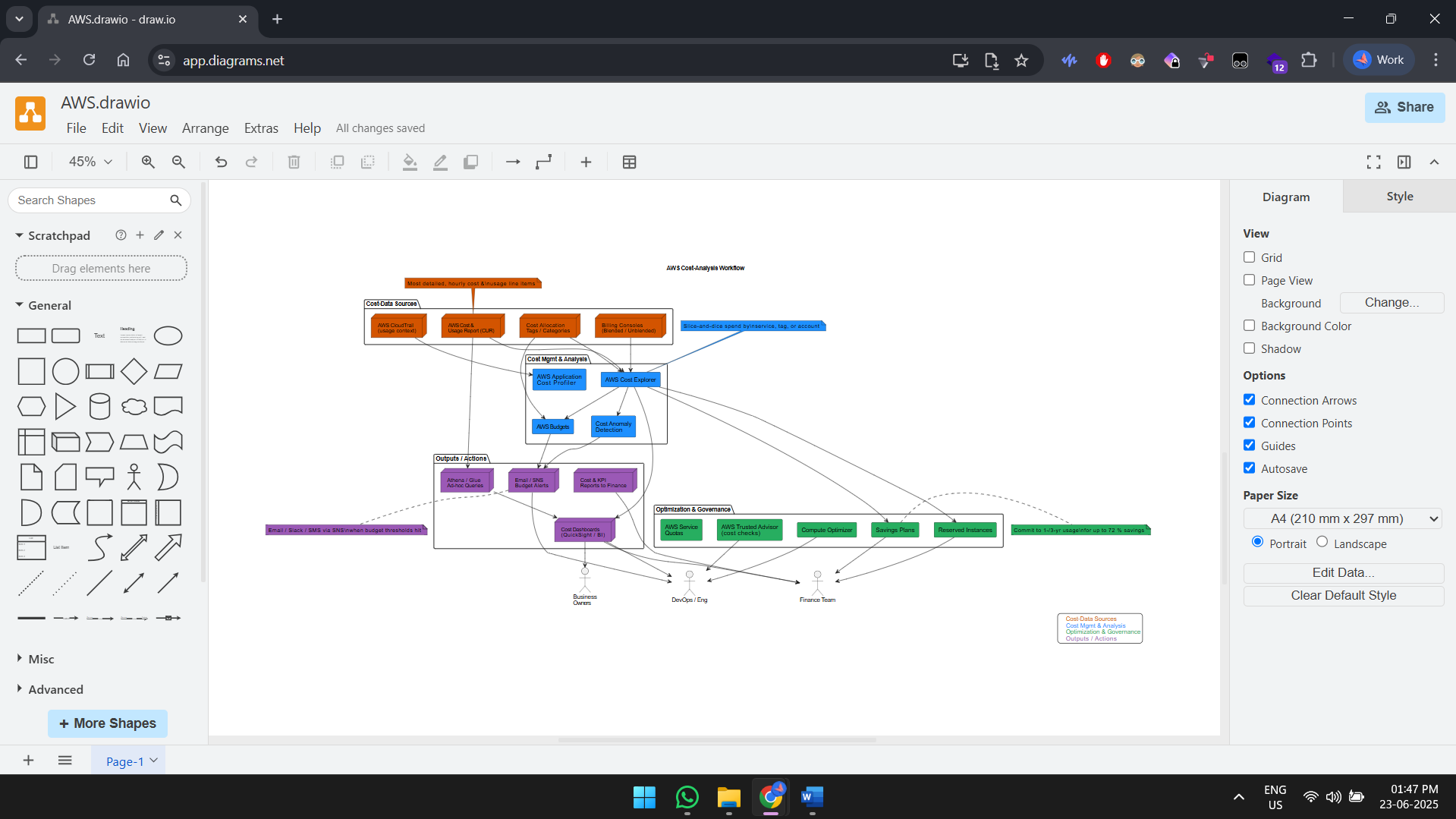
[AWS EBS 8](#_Toc201579752)

# AWS Deployment and Service Models



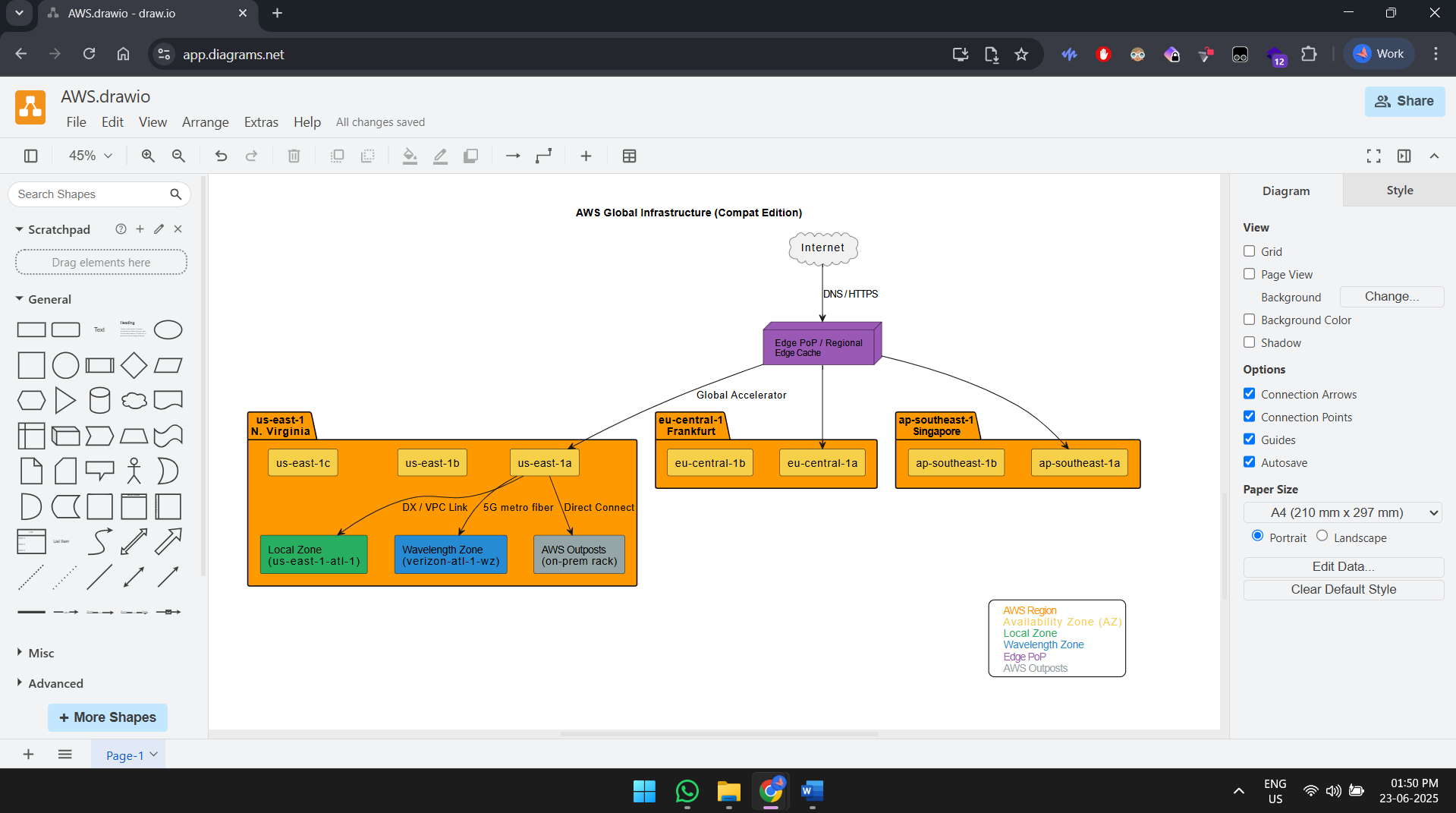
The diagram provides a graphical summary of the fundamental ideas of AWS cloud architecture, or rather helps further integrate Deployment Models and Service Models into one picture. Deployment Models (Public, Private Cloud, Hybrid, Community Clouds) determine the way and the place where AWS infrastructure is positioned, whereas Service Models (IaaS, PaaS, SaaS) specify the degree of control and abstraction that AWS will provide to the users. The diagram relies on color designation and AWS service imagery to differentiate every model and their common application scenarios. Dependencies among these layers emphasize flexibility of the approach to cloud adoption where an organization may opt to implement hybrid implementation with various levels of services depending on scalability, control, and cost-effectiveness requirements. Thes diagram is a high-level guide to IT architectures, cloud engineers, decision-makers designing AWS solutions.

# AWS Cost Analysis



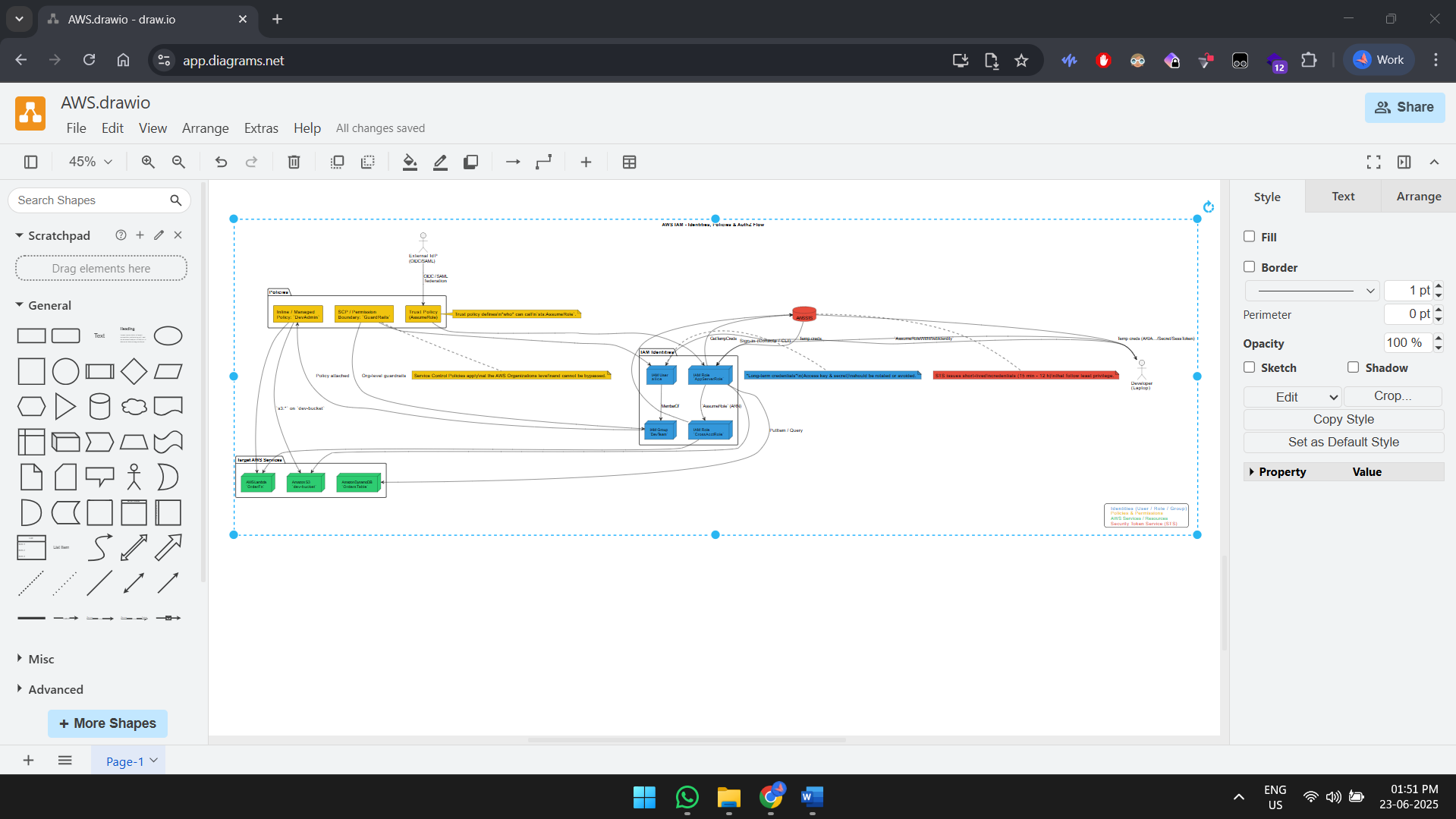
The Cost & Usage Report (CUR) is the foundation of cost analysis on AWS, providing line-item billing records you can load into Athena or Redshift to do deep dives; an upper layer on top of that consists of Cost Explorer, AWS Budgets, and Cost Anomaly Detection, to provide you interactive charts, threshold-based alerts, and machine-learning drifts warning. Assigning cost-allocation tags and AWS Organization cost categories, you split spend by team, project, or environment and bring charge-back or show-back reports to the finance. Usage patterns are translated into specific saving steps by optimization tools, Savings Plans, Reserved Instances, Compute Optimizer, and Trusted Advisor, and retention is automated by AWS Backup and Lifecycle Manager so the cost of storage does not surprise. The final outcome ends up being an assembly line: raw CUR is curated into dashboards by curation tools -> budget alarms are formulated in arcane budget alarms -> optimization levers are defined as a result of it, granular IAM permission secured, exportable to existing BI tools, or baked into the corporate ERP.

# AWS Global infrastructure



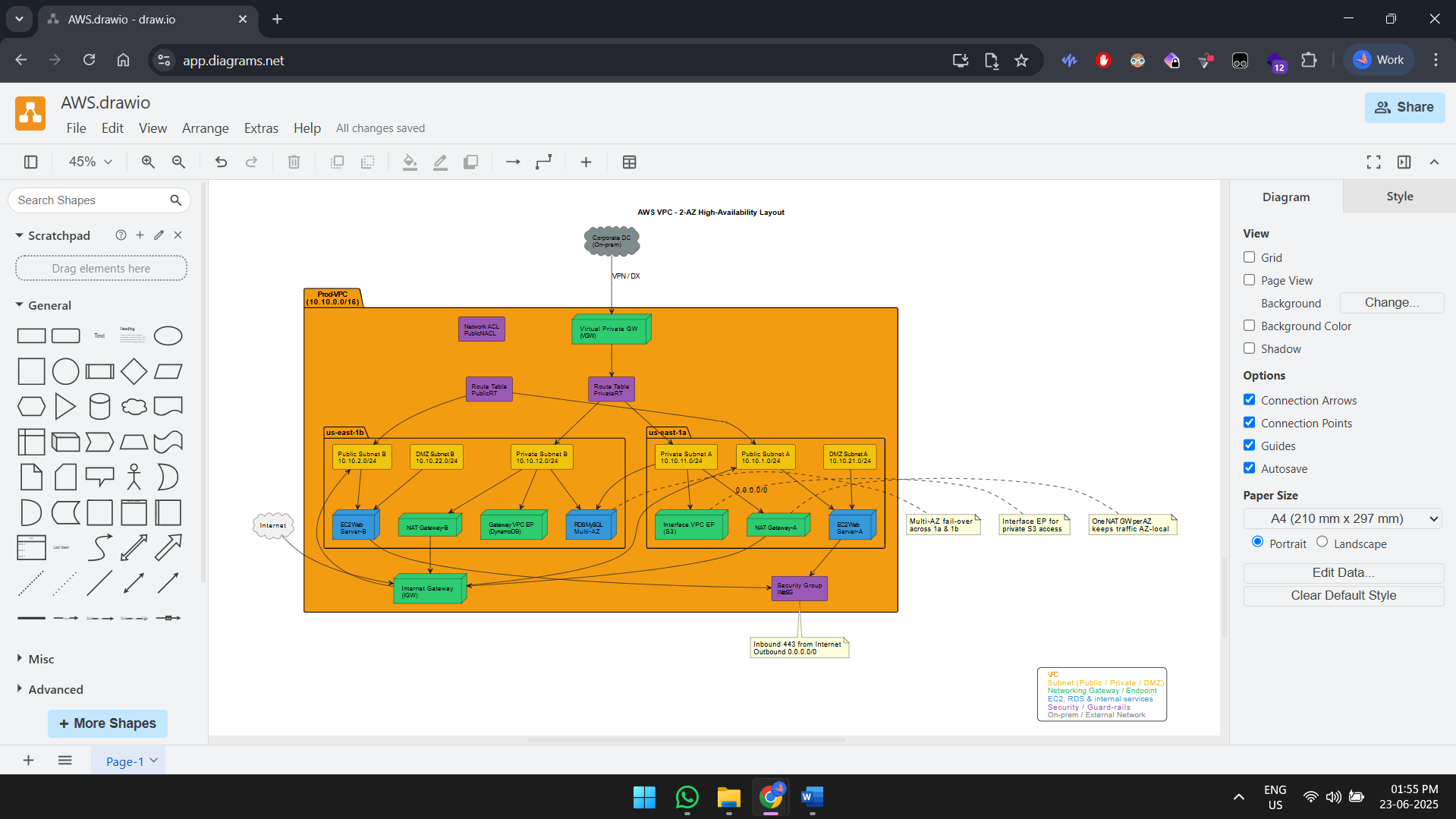
AWS has a layered worldwide presence: regions (consisting of at least three physically dissimilar Availability Zones) are the basis; local zones and Wavelength zones bring compute to the metropolitan sites and 5G carrier locations to achieve latency of sub-ten-milliseconds; there are 600 + edge locations and regional edge stores to accelerate CloudFront and Route 53 requests and Global Accelerator traffic; and Outposts or Snow devices integrate the cloud to customer premises. Each Availability Zone is separately powered, cooled and networked, and architects can build multi-AZ-tolerant workloads, whereas inter-region connections through the AWS backbone can deliver active-active architectures across multiple regions and disaster recovery. Sovereign and regulatory use cases are covered by compliance boundary (e.g. GovCloud, China regions, Top Secret regions). The ever-growing network supports simple S3 replication, all the way to globally distributed Aurora clusters and presents builders with a latency toresidency to resiliency spectrum.

# AWS IAM



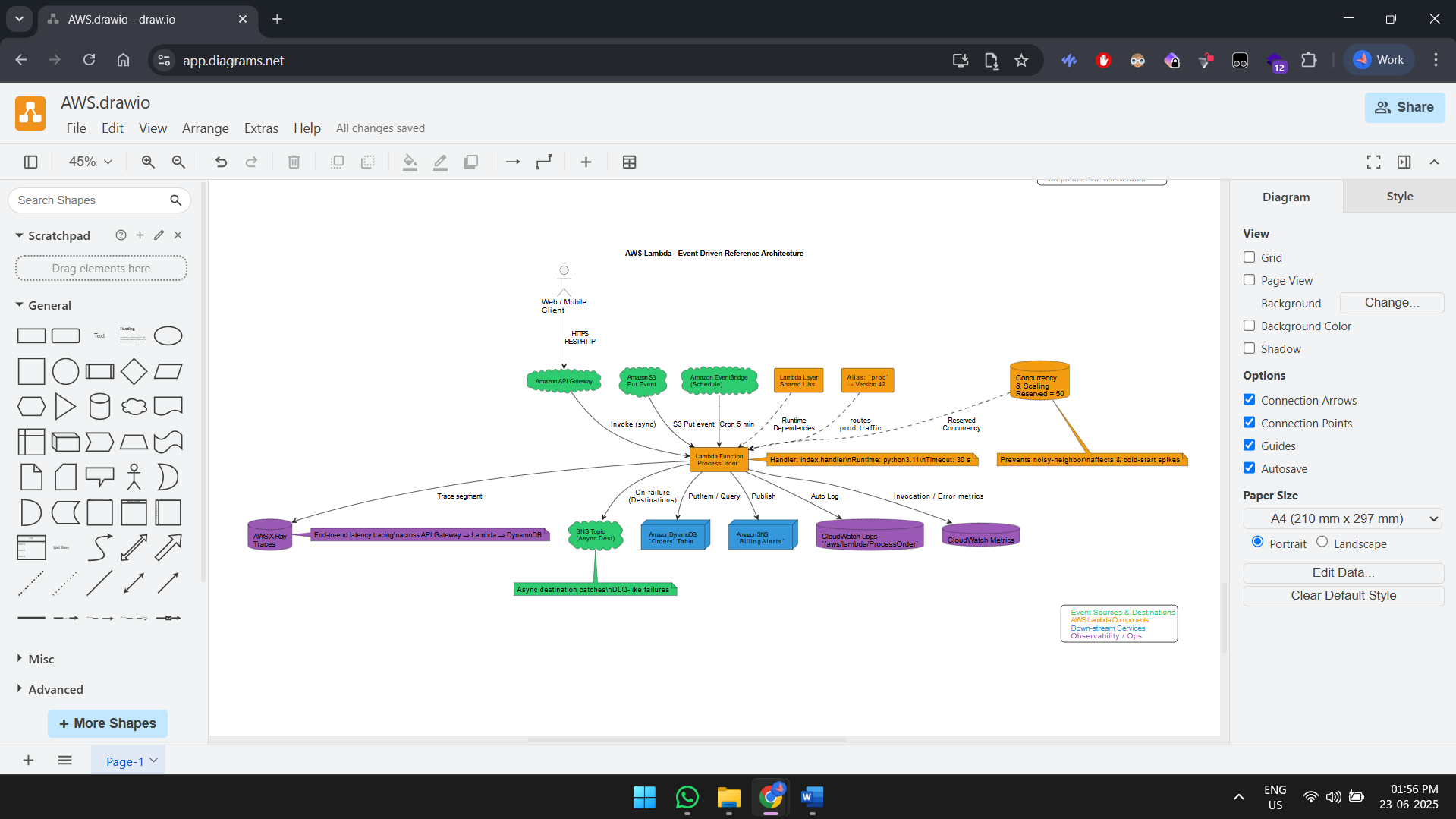
The control plane is IAM that specifies who is allowed to do what, which resource, and when. Users, roles, and groups are granted the permissions using managed or inline policies that are to be written in JSON; the evaluation logic denies explicitly with allowance thereafter, which is then narrowed by permission boundaries, session policies, and service-control policies in the AWS Organizations. Short lived credentials through the AWS STS promote scenarios in which roles are assumed within an account, and cross-account trusts, as well as federated identities with SAML/OIDC provider. By imposing fine-grained permits such as resource-level control, tag-based condition, and IAM Roles for Service Accounts (IRSA) to the EKS, the least-privilege design can be nailed down. Governance is also completed with audit trails in CloudTrail and the IAM Access Analyzer, which discloses unused permissions and the unintended methods of access by any other person.

# AWS VPC



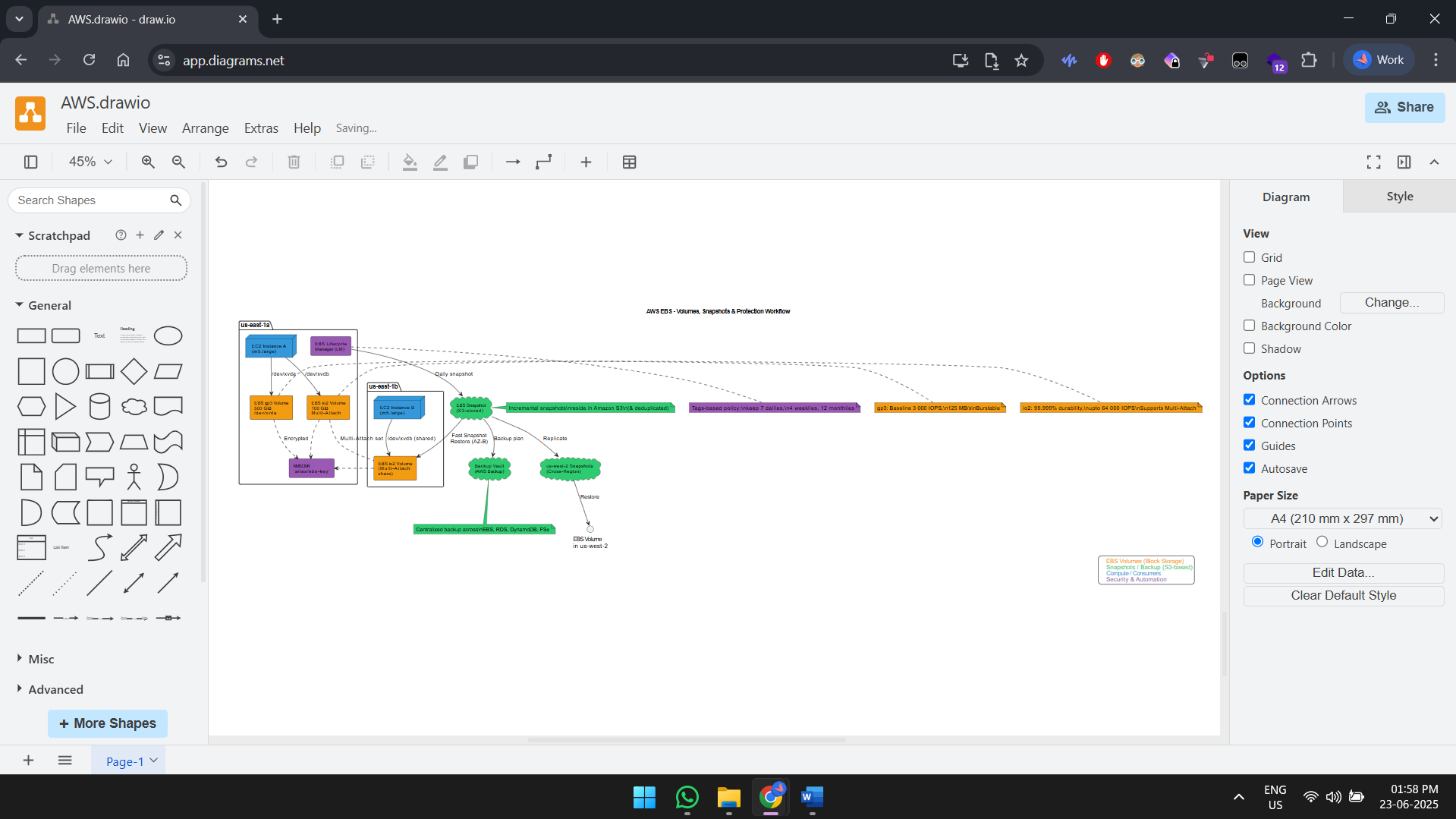
A VPC is a logically separated area of the AWS network in which you specify your own ranges of the IPv4/IPv6 address space, divide it into public, private, and isolated subnets in several Availability Zones, and hook up gateways: Internet Gateways, NAT Gateways, VPNs, Direct Connect, or Transit Gateway to manage ingress and egress. The route tables, stateless network ACLs, and stateful security groups control traffic flows, and allow you to realize either traditional three-tier or zero-trust topologies. VPC Endpoints leave S3, DynamoDB and most AWS services accessible without use of the public internet, and VPC Flow Logs can be fed into SIEM pipelines to threat hunt. It is possible to run network enable an all-software, enterprise-scale, with enterprise-scale capabilities built-in and Securtiy Group rules and prefix lists, such as the addition of Elastic Network Interfaces, features such as Reachability Analyzer and Network Firewall, making it possible to run enterprise-scale networks completely in prose.

# AWS Lambda



Lambda is the event-driven, fully managed compute service on AWS where you can upload the code (or even container images), select a runtime, and have the underlying platform handle provisioning, automatically scaling between thousands of concurrent executions, and automatically scaling to zero when not needed, all charging you only in milliseconds at a time. More than 200 native event sources (API Gateway, S3, EventBridge, DynamoDB streams, SQS, Kafka, Step Functions etc.) invoke Functions which can be wrapped by Versions to provide immutability, Layers to provide shared libraries and Aliases to provide safe blue/green or canary deployments. Such capabilities as Provisioned Concurrency and SnapStart calm cold-start latency, and concurrency quotas on a per-reservation and per-account allow avoiding noisy neighbor concerns. Logs and metrics are sent to the CloudWatch, traces are sent to X-Ray, and failures are sent to destinations or DLQs so that observability and resilient retries can be achieved without glue code. In tandem with IAM per-function roles Lambda allows teams to deliver micro-services or data-processing pipeline with a very small operational expenditure.

# AWS EBS



EBS also offers network-attached, AZ-scoped volumes of block storage, which appear as raw SSD/HDD devices to the EC2, container workloads, and certain Shared File Systems on premises, using special EBS Direct APIs. Types of volumes include low-cost HDD (st1/sc1), general-purpose SSD (gp3 with independent IOPS/throughput controls), high performance SSD (io2/io2 Block Express up to 256 k IOPS) and cold HDD, all of which support Elastic Volumes to live resize and change type. All the writes are asynchronously distributed in the AZ and replicated to guarantee 99.999% durability and can also be encrypted at rest KMS keys controlled by the customer. Incremental Snapshots store at S3, and can be restored very quickly, cross-AZ fast snapshot restore, cross Region disaster recovery, and also integrate with AWS Backup, Lifecycle Manager, and Recycle Bin.